My Name (myNetID)

IE598 MLF F18

Module 7 Homework (Random Forest)

Using the ccdefault dataset, and 10 fold cross validation described in Raschka;

**Part 1: Random forest estimators**

Fit a random forest model, try several different values for N\_estimators, report in-sample accuracies.

**Part 2: Random forest feature importance**

Display the individual feature importance of your best model in Part 1 above using the code presented in Chapter 4 on page 136. {importances=forest.feature\_importances\_ }

**Part 3: Conclusions**

Write a short paragraph summarizing your findings. Answer the following questions:

1. What is the relationship between n\_estimators, in-sample CV accuracy and computation time?
2. What is the optimal number of estimators for your forest?
3. Which features contribute the most importance in your model according to scikit-learn function?
4. What is feature importance and how is it calculated? (If you are not sure, refer to the Scikit-Learn.org documentation.)

**Part 4: Appendix**

Link to github repo